TEXT TO BRAILLE TRANSLATOR

NURAFNI BINTI HERMAN

A thesis submitted in partially fulfilment of the requirements for the award of degree

Of Bachelor of Computer Science (Software Engineering)

Faculty of Computer System & Software Engineering

Universiti Malaysia Pahang (UMP)

SEPTEMBER 2013

ABSTRACT

This thesis details the development of software to perform the translation of Text to Braille and support of translation tables to allow translation to Braille from many languages with the translation following grades rules. This application is improvements to the existing system in terms of features and abilities in translating the text into braille code. . Among those features that can be added to this application is to add the number of letters that can be translated. This application is a web based application using. This project uses use waterfall. The implementation of this application is using PHP, C# and MySQL as the database on a computer running windows 7.

CHAPTER 1

INTRODUCTION

This chapter briefly discuss on the overview of this project. It contains five sections. The section 1.1 is background of the project. The section 1.2 is the problem statement and motivation of this project. The section 1.3 is the aim and objectives where the project's goals are determined. The section 1.4 is scopes of the project. Section 1.5 describes about methodology for this projects. And Section 1.6 describe about the proposal organization which briefly describes the structure of this thesis.

1.1 Overview

Braille is a system developed by Louis Braille in the ninetieth century to allow the blind to read and write (Wikipedia, 2013). Charles Barbier originally developed this coded system, named 'Night writing' or sonography (American Foundation for the blind, 2009). In general, the code enabled soldiers to communicate silently and without light during the night (Wikipedia, 2013).

Louis Braille was inspired by this system and in turn then went to create a modified code to make more easy to use. Louis took several more years to make a system which was suitable to include all the letters of the French language. Eventually the Braille system has become widespread with the support of a group who was to become the Royal National Institute for the Blind (Wikipedia, 2006).

Braille code is a writing system using a series of raised dots to be read with the fingers by people blind or whose eyesight is not sufficient for reading printed material (The Unicode Standard, 2012). The standards describe about the characteristic of the braille code and the main features are briefly summarizes. Braille uses raised dots in groups of six which arrange in three rows of two (Blenkhorn, 1997). In generally, a group consisting of six dots which are arranged in three rows of two is known as the cell. (Blenkhorn, 1997). Each cell represents a letter, numeral of punctuation mark. Some frequently use words and letter combination also have their own single cell patterns. These six positions, which can be raised or flat, are used in combination to give just sixty four different braille "characters."

The earliest work on computerized translation of braille has been reported in a number of studies (Paul, 1997) such as: Association for Computing Machinery (1975), J M Ebersold(1985) and Peeters (1990). The use of context matching rules is much more straight forward and easier to understand. Consequently, the text to braille application has been constructed so that the bulk of the translation is achieved by using context specific rules.

1.2. Problem Statement

Reading is an easy task to the normal people who can read and write in a common language. However, as we know most of the blind people get very hard to face their life like other normal people. They are very difficult to read the common book as a normal people. They need to read book by using special book that use code that called braille code. Unfortunately, those books are difficult to get at a common bookstore. This is because lack of reading materials for blind people release by book publisher. Publishers fewer publish the book because of lack of manpower that can translate text to Braille code. So, a translator is needed to convert text message to Braille, so that a blind person can read it.

Nowadays, majority people in this world use a computer as the important technologies for communicate in their life after a mobile phone. These technologies help normal people in day life with different types of applications that can simplify work. One of the applications is text to braille code translator. However, most of this application is limited because consumers have to pay to use the full version. Most of these online facilities are half from the full version and can only translate fewer than 1,000 words only especially for open source system.

Based on the problem, a new solution is needed. The solution that can be done is by performing improvements to the existing system in terms of features and abilities in translating the text into braille code. Among those features that can be added to this application is to add the number of letters that can be translated. With these improvements, the number of characters that can be translated by the application is 1000 letters at the same time.

1.3. Aim and Objective

The aim of this project is to implement web-based application for translating text to braille code. The overall objectives of the project are:

1. To identify possible technique to converting text to braille translator based on specific code format.
2. To design the web-based application for translating the text to braille code by segmenting the image into Braille line
3. To implement the web based applications for Braille code translator

1.4 Project Scope

The scopes of this project are as follows:

1. The system can only translate 1000 words in one time.
2. The target users of this system are that beginner learner who is just learning the braille code.
3. Rules based technique as techniques used because of the pattern and condition for braille code.

Text to braille code translator is a project based on web-based application development. This application used grade one and grade two's braille code. Braille Grade one consists of 26 standards latter of alphabet and punctuation while braille grade two, a cell represent a shortness form of word. The combinations of this two braille grade are the best because this is the most popular of the grades of braille.

This application also allows user to translate text to braille code up to 1000 words at one time which is an improvement from previous application. The previous application only can translate less than 1000 word at one time. The targets users of this application are authors and publishers who want to translate from text to braille code.

1.5 Methodology

Methodology can properly refer to the theoretical analysis of the methods appropriate to a field of study or to the body of methods and principles particular to a branch of knowledge. There are five phase in methodology which is requirement, design, implementation, verification, maintenance.

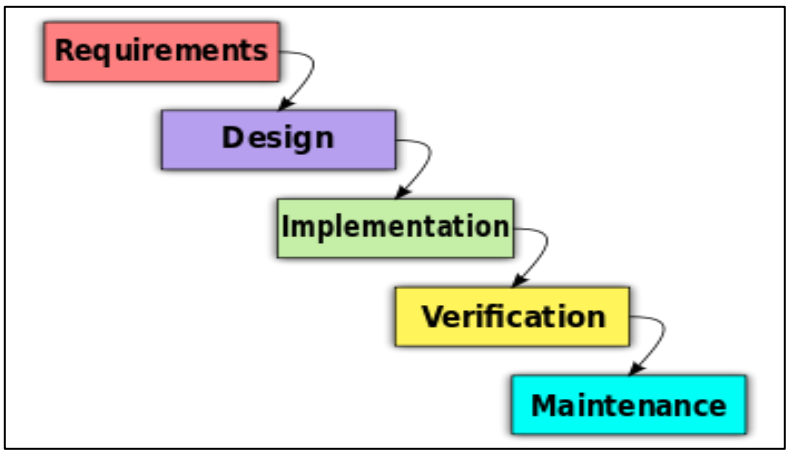

Figure 1.1: Methodology Phase

Requirement Analysis and Software Definition is the first phase of methodology which understands the requirements. The software definition must be detailed and accurate with no ambiguities. The activities are done by this phase will carry out problem statement, aim and objective, and scope of the project.

System Design is the requirements are broken down into logical modules for the ease of implementation. Hardware and software requirements for every module are identified and designed accordingly. Also the inter relation between the various logical modules is established at this stage. Algorithms and diagrams defining the scope and objective of each logical model are developed.

In short, this phase lays a fundamental for actual programming and implementation. This phase will be performed in chapter 2, Literature Review. In Literature Review, hardware and software that will used to develop this system will be identified. This phase also cover for chapter 4, Design.

System Implementation is the software process in which actual coding takes place. A software program is written based upon the algorithm designed in the system design phase. A piece of code is written for every module and checked for the output. This phase will cover in chapter 5, which will describe for next semester.

System Testing is programmatically implemented software module is tested for the correct output. Bugs, errors are removed at this stage. In the process of software testing, a series of tests and test cases are performed to check the module for bugs, faults and other errors. Erroneous codes are rewritten and tested again until desired output is achieved. This phase will cover in chapter 6, which will describe for next semester.

System Deployment and Maintenance is the final phase which the completed software product is handed over to the client after alpha, beta testing. The activities in this phase will be done by next semesters after all phase finish. This phase will cover for chapter 6.

1.6 Thesis Organizations

This thesis consists of seven (7) chapters. Chapter 1 will discuss about general background of the research focus, problem statement, research aim and the objective, scope, study module and thesis organization.

Chapter 2 is about literature review. First point of this is an introduction to highlight the points discussed on this chapter. This chapter contain research information that related to the project to get general overview about the techniques, application and approach by doing surveying.

Chapter 3 will explain on system methodology. It defines on methodology that used to develop the system and project planning. On this chapter also identify the necessary of the project such as software or tools on developing the system

Chapter 4 will elaborates on project design. This phases will discuss about flow of the project and interface for this system will design in this phases. On this chapter, the program determines either working properly or not.

Chapter 5 will elaborates on project implementation. Coding and testing for whole developments are implemented for this chapter. On this chapter, the program determines either working properly or not.

Chapter 6 describe on result and discussion that obtained from data analysis, project constraint and future work suggestion. Constraints are divided into two sections, development constraint and system constraint.

Chapter 6 this specific chapter would then summarize the result related to the real world problem and recommend some suggestion for final product.

CHAPTER 2

LITERATURE REVIEW

The purpose of this chapter is to present a selected literature review, which is very important for the research. This chapter also describes and explains on the literature review carried out on the system. Besides that, previous research also will be discussed in this section at least one existing system and techniques that being used in other research which is related to this system will be explained and compared to highlight the differences.

2.1 Braille System

2.1.1 Braille Code

Text to braille translator is an application to be developed to translate the text to braille code. This application comprises of a cell per character and themselves consist of six raised dots arranged as shown below. The ways of dots are raised gives meaning and each cell has the meaning of one letter, number or simple punctuation mark.

The numbering system for the dots in the Braille cell is shown in Figure 2.1.It shows that the Braille cell contains six dots which are formed by two rows of three. For example, Figure 2.2 has the dots 1, 2 and 5 raised which give the letter 'h'. These six positions which can be raised or flat, are used in combination to give just 64 different braille "characters."

This clearly means that there can't be a one-to-one correspondence between braille characters and texts. In the simplest commonly used form, called Grade 1 braille, the lower case letters A-Z and the major punctuation symbols are represented by a single braille characters, with "shift" character being used to indicate other information such as upper case, digits, and italics.

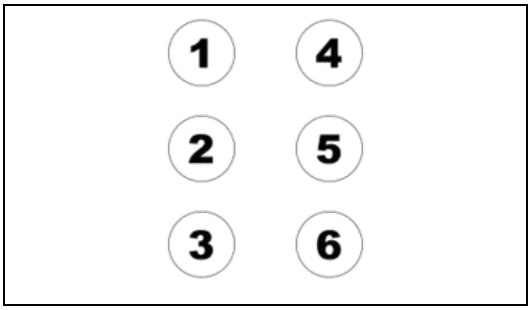

Figure 2.1 Braille Cell Numbering

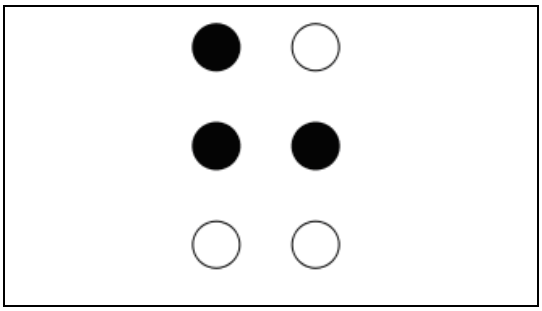

Figure 2.2 Braille Cell representing 'h'

2.1.1 ASCII Braille

ASCII Braille is a set of numerical characters codes to representing six-dot braille cells electronically. ASCII Braille formally known as The North American ASCII code. Braille ASCII is a subset of ASCII characters set and consists of 64 of the printable ASCII characters to represent the combination in six –dots Braille. ASCII Braille is also known as Grade zero Braille.

There are many different ASCII braille codes with various assignments of characters codes to braille cells although the one produced by the Duxbury Braille Translator (DBT), which is sometimes called North American or MIT ASCII Braille, probably the most extensively used. Two of the file formats that are used North American ASCII braille are Braille Formatted (BRF) files and Duxbury Translator Braille (DTB) files.

This system would take characters from the ASCII tables of values 32 to 95 inclusive and map these to the Braille system. Therefore, this system has a separate set of Braille cells reserved for capital letters, lower case letters and numbers. The full of the ASCII characters is shown in Figure 2.3.

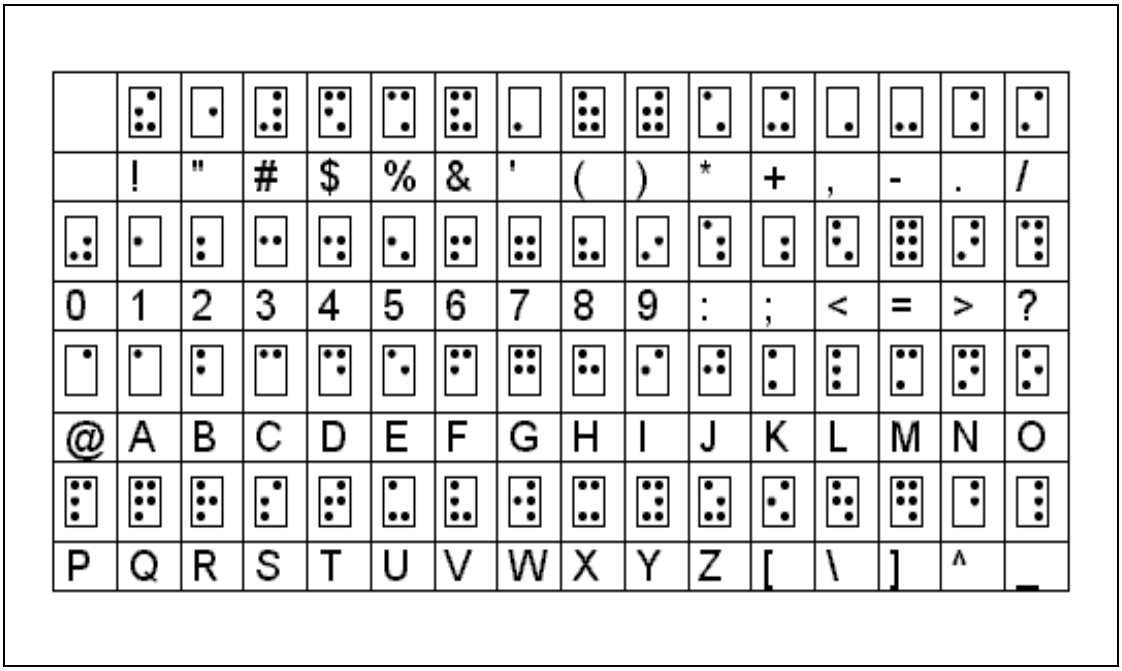

Figure 2.3 North American ASCII Braille Table

2.1.2 Braille Grade

Braille Grade is a number of different versions of Braille. There are three types version which is grade one, grade two and grade three. First grade in braille grade is grade one. Grade one consists of 26 standard letters of alphabet and punctuation. Figure 2.4 show table for Braille Grade one. In grade one, each possible arrangements of dots within a cell represents only one letter, number, punctuation sign, or special braille composition sign which is a one-to-one conversion.

Individual cells can't represent words or abbreviations in grade one. Because of this grade's inability to curtail the words, books and other documents produced in grade one's braille are bulkier and large than normally printed text. Grade one only used for beginner learner to read Braille. But in early 2000s a new movement was in place among elementary school teachers of braille to introduce children with sight difficulties to grade two's braille immediately after teaching basic grade one's braille.

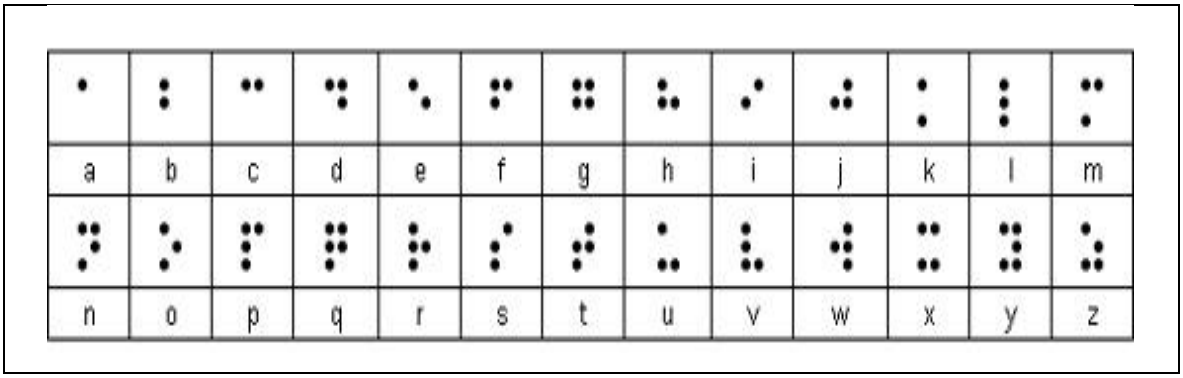

Figure 2.4 Grade one's Braille table

Grade two's braille was introduced as an alternative of space-saving to grade one's braille. In grade two's braille, a cell can represents a shortened form of a word. Many cell combinations have been created to represent common words, making this grade is the most popular of the grades of braille. There are part-word contractions, which often stand in for common suffixes or prefixes, and whole-word contractions, in which a single cell represents an entire commonly used word.

Words may be abbreviated by using a single letter to represent the entire word, using a special symbol to precede either the first or last letter of the word while truncating the rest of the word, using a double-letter contraction such as "bb" or "cc", or removing most or all of the vowels in a word in order to shorten it. Figure 2.5 show list of latter for braille grade 2. A complex system of styles, rules, and usage has been developed for this grade of braille.

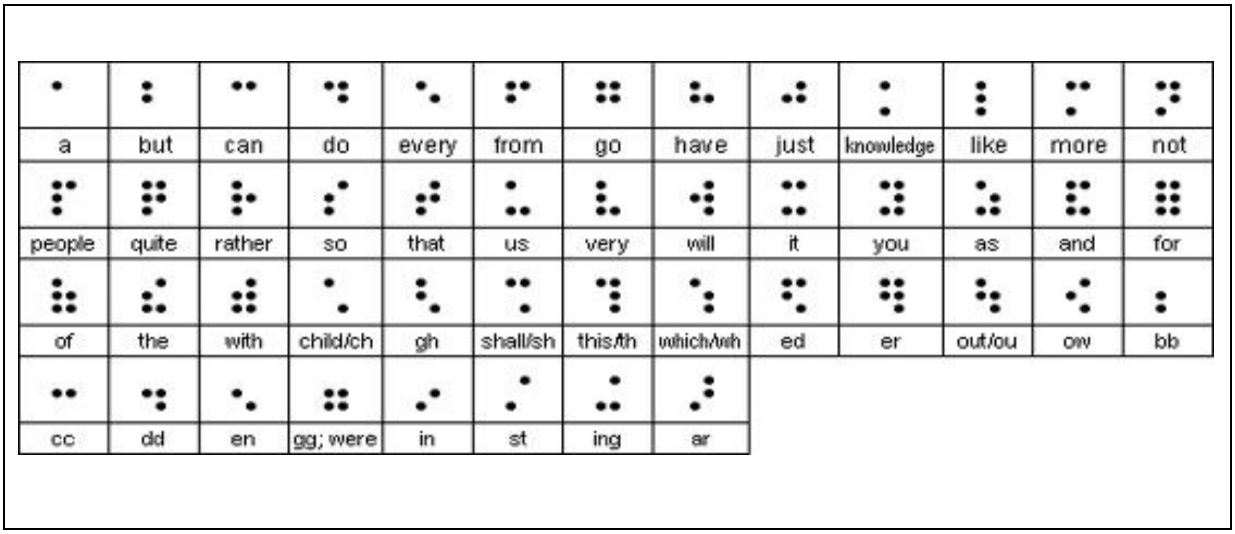

Figure 2.5 Grade two's Braille table

Grade three is the last of the grades of braille which essentially a system of braille shorthand. Because it has not been standardized, it is not used in publications. Instead, it is typically used by individuals for their own personal

convenience. It contains over 300 word contractions and makes great use of vowel omission. In addition, the amount of spacing between words and paragraphs is decreased in order to shorten the length of the final document. It also sometimes substitutes combinations of punctuation symbols for words. For a list of grade three symbols

2.2 Text to Braille Translator

2.2.1 German contracted braille translation

The early work on computerized translation of braille was basically concern with translation from text in braille. One solution for instance is the use of production rules derived from a Markov system. This approach has been followed by W.A Slaby. However, this solution results in a rapid increase of the number of production rules.

Slaby developed another system for German contracted braille translation in 1980. This system allows non-expert to modify rules to perform translation of different language into braille. This algorithm also applicable to the braille to text translation besides can use by text to braille translation. In 1995, a system to convert standard English to braille into text proposed by Paul Blenkhorn base on Slaby' system. This method uses a decision table with input classes and states and rule table containing all rules for containing all rules for translation. People who are not expert in computer algorithm can be readily updated and modified for this system.

All rules are listed in ASCII alphabetical rule according to Blenkhorn's algorithm. For rules whose focuses start with the same character(s), the order in

which they appear in the table is related to their priority. So the rules have to be checked in order from top to bottom. The first rule which is found has been used. Therefore, actually the Blenkhorn's algorithm can be regarded as a Markov system.

In Blenkhorn's system, there are totally 498 rules in translation table. Because the format of each rule is simple and each part of one rule can be checked separately, therefore, it takes the system easily be implemented in hardware.

2.2.2 Braille Translation By Lee

The Braille Translation Program (Lee, 2006) is a program for Apple Mac OS X that translates the text to Braille as it is typed into the editor, and allows for the text to be edited. Changes are reflected immediately in both the ASCII Braille and Braille output. It supports translation into both Grade One and Grade Two Braille in different languages.

The Braille Translation Program also supports the translation of Braille into print besides supporting the translation of print into Braille, it. The translation algorithm used was developed by Shannon Thrower (Thrower, 2005). Figure 2.5 shows a screenshot of the Braille Translation Program, with a section of text translated into Grade One Braille.

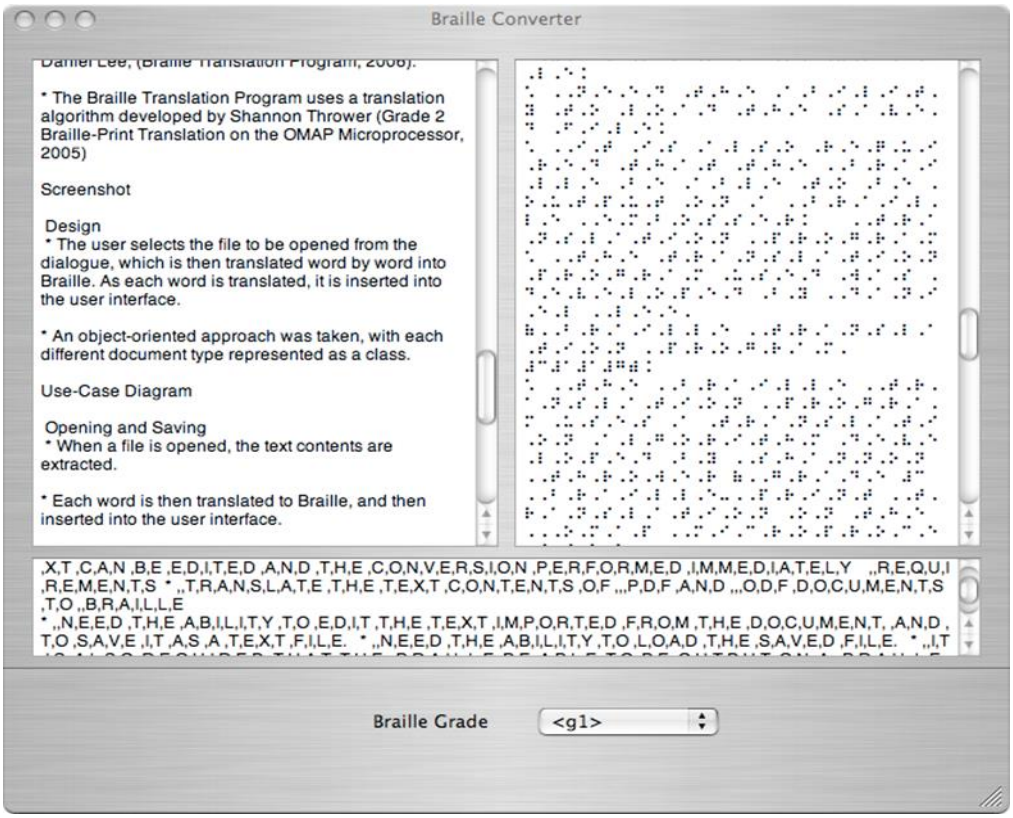

Figure 2.6 Screenshot of Braille Translation Program

2.2.3 Complexities of Braille

In generally in braille system one character can have many meanings, depending on where it is placed within a word and to which grade it was translated into.

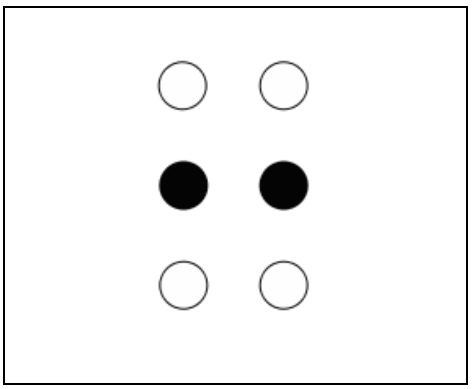

Figure 2.7 Braille Cell representing '3' in ASCII

For example, the character with dots two and five raised has three meanings. For grade two, it could mean ‘cc’ if it is in the middle, ‘con’ if it is at the front of a word or could mean the number three in the ASCII Braille. Another example would be the character with the dots two, three and five raised is the letter ‘f’. This character could mean ‘ff’ when used in the middle of a word, at the beginning of a word, it would mean ‘to’ and at the end of a word, the same character represents an exclamation point.

The contraction rules take into account the linguistic structure of the word; thus, contractions are not to be used when their use would alter the usual Braille form of a base word to which a prefix or suffix has been added. And some portions of the transcription rules are not fully codified and rely on the judgement of the transcribe. Thus, when the contraction rules permit the same word in more than one way, preference is given to "the contraction that more nearly approximates correct pronunciation." (Wikipedia, 2006).

Another example of this complexity is shown where ‘wh’ can be replaced by one character, except when the word is actually two words combined. So the ‘wh’ in a while can be replaced by a character, but the ‘wh’ in rawhide cannot be replaced by one character. Trying to find out whether or not a word is actually two words combined together would require a large dictionary and thus would also an increase the CPU usage with very little benefit in the actual translation since this occurrence would occur infrequently at best.

Thus, for computerized Braille translation, either both rules would either be left out of the translation tables or allow the contraction into one character. Also, contractions could not be used across syllable boundaries, would fall into the above category. In Table 2-1 (Shannon, 2005) shows the relationship between Print, Braille and Computer Braille representations of the word ‘knowledgeable’.

This shows how parts of the word are contracted to different characters. For example the part of the word ‘know’ is translated to the cell one-three which is represented in ASCII by the character ‘k’, ‘ed’ is contracted to the cell one-two-four-six which in ASCII is represented by the ASCII character “$” with the letters ‘ble’ being contracted into the ASCII character ‘#’ which is the Braille cell of three-four-five-six.

Print		knowledgeable
Grade 2 Translation	Braille Characters	
	ASCII Characters	KL$GEA#

Table 2.1: Relationship between Print, Grade 2 Braille and Computer Braille

There is also the extra problem with grade two Braille is that in translating into this grade gives problems with loss of translation. For example, translating | and / into Braille gives the same character, thus losing the capability to translate back into English.

There are also problems as different cultures or people using different languages would use different Braille translation rules which can cause confusion. So having a system where Braille can be translated into one language and then into Braille system of a different nature so that the same text can be translated from one Braille system into another would be useful.

2.3. Translator Technique

Techniques that implement in text to braille translator are rules-based technique. Rules-based technique is one of artificial intelligence technique. Rule-based methods, rule discovery or rule extraction from data, are data mining techniques aimed at understanding data structures, providing comprehensible description instead of only black-box prediction. Rule based systems should expose in a comprehensible way knowledge hidden in data, providing logical justification for drawing conclusions.

Rule based also showing possible inconsistencies and avoiding unpredictable conclusions that black box predictors may generate in untypical situations. Sets of rules are useful if rules are not too numerous, comprehensible, and have sufficiently high accuracy. Rules are used to support decision making in classification (Classification, Machine Learning), regression (Regression, Statistics) and association tasks.

Various forms of rules that allow expression of different types of knowledge are used. In text to braille translator, there is a set of rules that should be in use. This set of rules are not too numerous, comprehensible, and have sufficiently high accuracy. This set of rules is important to make the decision to get the correct code.

2.4 Development Tools

To develop Text to Braille translator, it is important to choose the appropriate software and hardware. The following are the hardware and software to develop Text to Braille translator. The Specifications of hardware and software to develop applications is shown in Table 2.1 and Table 2.2.

a) Software

No	Software	Development
1.	Windows 7 Home Premium	As the operating system
2.	Apache XAMPP	Database for store data
3.	Macromedia Dreamweaver	To design interface and development system
4.	Microsoft Office Word 2010	Documentation and report making
5.	Microsoft PowerPoint 2010	To do system presentation
6.	IBM Rational Rose	Create Use Case
7.	Microsoft Office Project 2010	To produce system Gantt Chart
8.	Adobe Acrobat Reader 9.0	To read internet resources

Table 2.2: Software Requirement

b) Hardware

No	Item	Quantity	Purpose
1	Dell Inspiron 5420	1	-Prepare proposal and document -Design and develop the system
2	Pen drive	1	Data transfer includes documentation and system
3	Hard drive disk	1	-Data transfer includes documentation and system -For backup data
4	Printer/Scanner	1	Print documentation

Table 2.3: Hardware Requirement

2.4.1 Operating System

An operating system, or OS, is a software program that enables the computer hardware to communicate and operate with the computer software. Without a computer operating system, a computer would be useless. It manages all programs in a computer.

The most popular operating systems in computer field are Linux, Windows, and Macintosh. They are undoubtedly the major operating systems used by computer users around the world. Windows has the highest user base, followed by Linux and Mac. For develop Text to Braille translator, Windows has been chosen.

Windows is a personal computer operating system from Microsoft that together with some commonly used business applications such as Microsoft Word and Excel. The Windows operating system is pretty versatile. It can be installed on PCs having variable amounts of resources. Windows users get best choices for their operating systems based on the system resources they have. It cannot be installed in Macintosh computers.

Benefit of Windows users is the user-friendliness of the operating system. The graphical user interface is surprisingly easy to work with, and many complicated tasks on Windows can be performed with a few clicks of a mouse only. Windows has the highest user correspondingly and most software vendors develop software packages, tools and utilities based on this operating system (Windows 7, 2009).

2.4.2 XAMPP

XAMPP is a free and open source cross-platform web server solution stack package, consisting mainly of the Apache HTTP Server, MySQL database, and interpreters for scripts written in the PHP and Perl languages. It is a free web server capable of serving dynamic web pages.

XAMPP was intended as a development tool to allow website designers and programmers to test works on own computer without any access to internet.The special tool is provided to password protected. XAMPP also provides support for creating and manipulating databases in MySQL and SQLite among others.

Once XAMPP is installed, it is possible to treat a localhost like a remote host by connecting using an FTP client. It is also possible to connect to localhost via FTP with a HTML editor. The default FTP user is "new user", the default FTP password is "wampp".The default MySQL user is "root" while there is no default MySQL password (Oswald et al, 2002-2012).

2.4.3 PHP

PHP stand for PHP: Hypertext Pre-processor. It is a powerful tool for making dynamic and interactive web pages. It is the widely used, free and efficient alternative to competitors such as Microsoft's ASP. PHP is a popular language for server-side applications.

It has evolved to include a command line interface capability and can be used in standalone graphical applications. It is a widely-used general purpose scripting language that is especially suited for web development and can be embedded into